International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 251-278
https://doi.org/10.70917/ijcisim-2026-3582
Article

NSAR-T: A Fine-Grained Aspect-Based Sentiment and Transformer Framework for Personalized Recommendations

Sonal Gupta[1*], Indu Kashyap[2]

[1*]Department of Computer Science and Engineering Manav Rachna, International Institute of Re-search Studies, sonal.gupta2080@gmail.com
[2]Department of Computer Science and Engineering, Manav Rachna, International Institute of Re-search Studies, indu.set@mriu.edu.in

Abstract: The rapid growth of digital platforms and online review systems has significantly increased the need for intelligent recommendation systems capable of understanding user preferences and contextual sentiments. However, traditional recommendation models often fail to capture fine-grained emotional variations, aspect-level opinions, and semantic relationships present in textual reviews, leading to reduced personalization accuracy. This work proposes a Nuanced Sentiment-Aware Recommendation with Transformers (NSAR-T) model incorporating Aspect-Based Sentiment Analysis (ABSA) with the Robustly Optimized Bidirectional Encoder Representation from Transformers (RoBERTa) to enable the generation of more granular personalized recommendations. A framework for hotel recommendations based on a combination of data sources consisting of 200,000 scraped review records from the primary source of hotels, 878,561 TripAdvisor reviews, and 37.6 million records for Expedia recommendations. The methodology consists of data preparation, extraction of aspects, sentiment classification based on RoBERTa, seven varieties of fine-grained modelling of sentiment, aggregation of sentiment, Principal Component Analysis (PCA) based feature engineering, Cross-Validation (CV), and generation of recommendations from the Top-10 list. The results show an accuracy of 99.85%, precision of 99.91%, recall of 99.81%, F1-Score of 99.86%, Root Mean Square Error (RMSE) of 0.03, Mean Average Precision (MAP) of 0.99, and Normalized Discounted Cumulative Gain (NDCG)@10 of 1; demonstrating superiority over currently employed machine learning and transformer-based recommendation systems with regard to the relevance of recommendations, contextual knowledge of recommendations, and how personalized recommendations are created.
Keywords: Aspect-Based Sentiment Analysis (ABSA), Fine-Grained Sentiment, Personalized Recommendation, Recommendation Generation, Sentiment Aggregation

1. Introduction

Digital platforms, including e-commerce, online travel services, and social media platforms, have experienced rapid expansion in recent years, resulting in a substantial increase in the volume of accessible data [1]. The information overload has created difficulties for users who want to find products and services and content that match their interests [2]. Personalized recommendation systems now serve as essential elements for contemporary intelligent systems, which solve this problem. The system analyzes user behavior through their past interactions and preferences to deliver personalized item recommendations that match their interests [3]. The practice of providing personalized recommendations improves user experience while enhancing user engagement and retention and decision-making processes across different fields, as shown in Figure 1(Figure 1: Applications of Recommendation Systems).

Figure 1: Applications of Recommendation Systems [4]

1.1 Evolution of Recommendation Systems

Recommendation systems have evolved through continuous development work from their original heuristic methods to their current advanced data-driven systems [5-7]. The early development phase of recommendation systems was dominated by traditional methods, which included matrix factorisation, content-based filtering, and collaborative filtering [8-10]. Content-based filtering recommends products based on product attributes and user profiles, while collaborative filtering predicts consumer preferences through user actions from similar users [11,12]. Matrix factorization techniques further improved these methods by learning latent representations of users and items [13,14]. The approaches have gained popularity because they enable users to manage structured data with high-performance results [15].

Current recommendation systems show insufficient ability to understand sentiment and semantic nuances because they fail to handle complex linguistic phenomena, which include sarcasm and implicit opinions and different levels of sentiment intensity [16]. Data sparsity and cold-start problems still lack effective solutions because limited user-item interactions make it difficult to create accurate recommendations [17,18]. Existing models face a major problem because they fail to adjust dynamically to user preferences that change in real time. The scalability problem exists because modern recommendation systems need to manage millions of users and items in large-scale data environments [19,20]. Figure 2 below shows the evolution of recommendation systems (Figure 2: Evolution of recommendation systems (Source: Author's Own)).

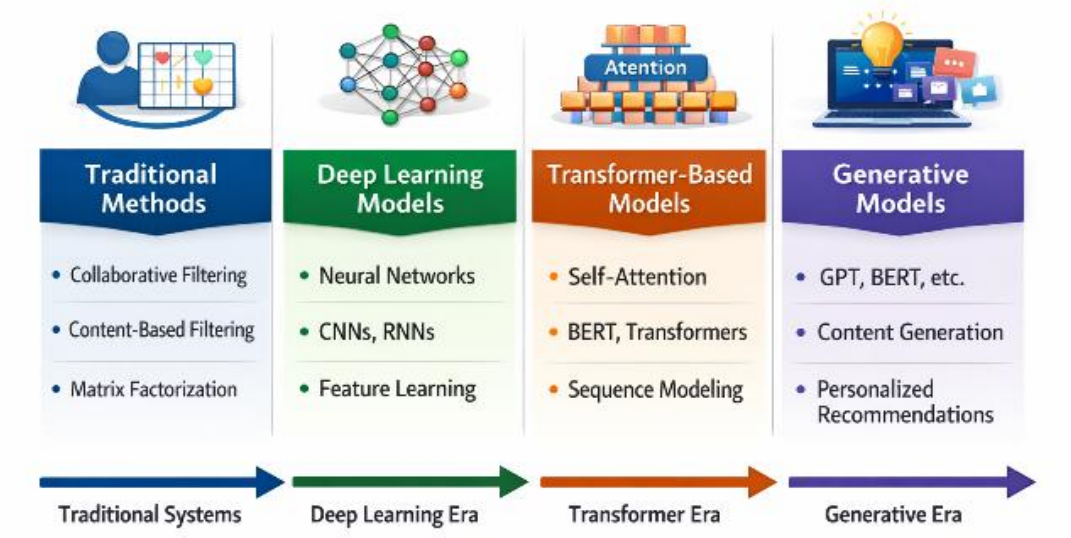

Figure 2: Evolution of recommendation systems (Source: Author's Own)

In recent times, transformer-based recommendation systems have emerged as the preferred choice because they excel at understanding contextual and semantic relationships between sequential and textual data elements [21-24]. The models, such as Bidirectional Encoder Representations from Transformers (BERT) and RoBERTa, use self-attention mechanisms to obtain user-generated textual review data, which helps them understand user

preferences more effectively [25-27]. Transformer models demand extensive computational power together with high processing requirements to operate successfully in large-scale environments [28,29]. The field of generative recommendation models include Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), which researchers have used to create artificial user-item interactions that enhance recommendation system reliability [30-32]. The models assist researchers in solving sparse interaction data problems, yet they require sophisticated system designs that result in unpredictable training results [33].

1.2 Fine-Grained and Nuanced Sentiments

Fine-grained sentiment refers to the process of classifying user opinions beyond traditional positive, negative, and neutral categories by identifying subtle variations in emotional intensity and contextual meaning [34,35]. Nuanced sentiments capture detailed emotional expressions such as mild satisfaction, strong approval, slight dissatisfaction, sarcasm, and mixed opinions that are commonly present in user reviews [36]. Fine-grained sentiment modelling enables users to gain better insights into how users express their preferences than coarse-grained sentiment analysis.

In the current work, the sentiments are classified into seven levels of sentiment to identify sentiment strength and emotions appropriately. Sentiment classification facilitates the identification of semantically dissimilar sentiments and thus enhances personalized output generation. The suggested approach deals with the identified sentiments using ABSA, whereby reviews are segmented into aspects including service, cleanliness, location, and price, with sentiments being analyzed for each individual aspect.

1.3 Role of Transformer Models in Handling Fine-Grained Sentiments

Transformer-based language models enable accurate analysis of complex emotional expressions through their advanced contextual learning capabilities. The BERT and RoBERTa models employ self-attention to detect dependencies and relations between different words in text reviews [37,38]. The model achieves comprehension of complex language structures and hidden emotional meanings through this process.

RoBERTa is used in the proposed system for ASBA because it has a more optimized training process and more robust context understanding than regular BERT. RoBERTa performs very well when it comes to recognizing the sentiments that belong to certain aspects through semantic link identification in reviews and aspects that have been mined from the dataset. Semantic feature representation through transformer embeddings allows for more accurate fine-grained classification. The model also improves recommendation quality through this enhanced semantic understanding.

Specifically, the objectives of this study are:

- To develop a recommendation framework that integrates ABSA and transformer-based modelling for improved personalization.
- To incorporate fine-grained sentiment modelling with continuous sentiment representation for capturing nuanced user preferences across multiple aspects.
- To perform aspect-level sentiment classification for identifying sentiment polarity across different review attributes.
- To evaluate the effectiveness of the proposed framework using standard performance metrics such as accuracy, precision, recall, and F1-score.

1.4 Novelty and Contribution of the Study

The proposed NSAR-T model represents the innovative element of this research study. The system combines ABSA with RoBERTa-based transformer modeling to create a single operational system. The NSAR-T model introduces a sentiment analysis system that enables users to categorize their feelings into seven different emotional states. It provides a comprehensive emotional assessment and contextual emotional strength assessment that goes beyond standard sentiment analysis methods. The proposed NSAR-T framework creates a new method for sentiment-based recommendations through its combination of aspect-level sentiment extraction, transformer-based semantic understanding, and weighted sentiment aggregation, which enhances the accuracy of personalization, the relevance of recommendations, and the ability to understand personalized recommendation systems.

The main contributions of this paper are summarized as follows:

- Development of the NSAR-T framework, integrating ABSA and RoBERTa into a unified transformer-based recommendation architecture.

- Integration of a fine-grained seven-level sentiment classification mechanism for capturing nuanced and context-aware user sentiments across multiple review aspects.
- Implementation of a weighted sentiment aggregation strategy for transforming aspect-level sentiment outputs into recommendation-oriented sentiment scores.
- Design of a sentiment-aware recommendation ranking mechanism that generates personalized recommendations purely based on aggregated fine-grained sentiment scores.
- Empirical evaluation of the proposed NSAR-T framework using standard performance metrics such as precision, recall, and F1-score.

The rest of the paper is organized as follows: Section 2 discusses the previous studies, and Section 3 includes the dataset used and the proposed methodology. Section 4 illustrates the results, and Section 5 concludes the research along with the future scope.

2. Related Work

In this section, a thorough discussion of the most up-to-date work in ABSA is presented below:

2.1 Deep Learning (DL) Based Recommendation Models

[39] introduced a Personalised Adaptive Multi-Product Recommendation System (PAMR) utilising transfer learning as well as Bidirectional Gated Recurrent Units (Bi-GRU). The accuracy for this model was 96.9%, and the values of precision and F1 score was 96.2% and 96.97%, respectively, using the Amazon data set. Nonetheless, the model only emphasized enhancing the accuracy of recommendations while not considering the extraction of aspects from customers' reviews on any given product.[40] employed a data-driven approach for developing a custom-made recommendation system for visitors of Nepal. A survey was conducted involving 2400 foreign as well as local tourists. There were 125 attributes in the questionnaire. Data had been pre-processed, and selected features were used to enhance the performance of the model. It was found that the proposed methodology outperformed previous systems and was better able to offer optimized recommendations to tourists visiting Pokhara. However, the study had limited use of text analytics and depended upon structured survey responses only. [41] developed a recommendation engine by combining the concept of sentiment analysis with the matrix factorization approach and deep learning algorithms. The proposed recommendation engine used supervised machine learning algorithms to perform sentiment analysis and extract implicit features. To improve scalability during recommendation engine design, the Singular Value Decomposition algorithm was adopted. It was revealed from the results that the proposed approach improved the efficiency of the Collaborative Filtering (CF) algorithm. Nonetheless, the process of sentiment analysis was quite coarse, and hence it could not detect fine-grained sentiments.

2.2 Sentiment Analysis in Recommendation Systems

[42] proposed MulMoSenT, an innovative approach to image-text sentiment alignment that addressed practical problems related to multimodal sentiment analysis, especially for low-resource languages. This model was empirically shown to have an accuracy rate of 84.90%, higher than all baselines. Nonetheless, this framework was mainly concerned with multimodal sentiment alignment but failed to discuss recommendation system integration and aspect-based opinion mining. [43] introduced the concept of the Gated Dynamic Local-Global Attention (GDLGA) model, which can enhance the accuracy of sentiment analysis on short messages while capturing both global context and sentiment clues locally. It proved better than traditional methods, attaining accuracy scores of 87.62% and an Area under the Curve (AUC) of 0.9484 using the Sentiment140 data set. While performing well, the method had to go through complicated feature extraction and consumed many computing resources.[44] conducted a sentiment analysis, topic modeling, and text mining on 2000 reviews of Expedia and Booking.com apps from the Google Play database to evaluate the consumers' emotions and decision-making behavior.

In the process of the study, Naïve Bayes (NB) was employed for classification purposes, while Python's WordNet was used for text mining, revealing that 81.9 percent of the Booking.com app reviews were positive, whereas only 55.8 percent of the Expedia ones were positive. For Booking and Expedia, the F1-score values were 0.977 and 0.933, respectively. Nevertheless, the study paid attention primarily to the polarity of the reviews and neglected the aspect-based sentiments. [45] developed a new hybrid approach for analysing consumer sentiment. In the preprocessing stage, the NLP process was used to eliminate unwanted data from the review text. The performance evaluation of the proposed framework was done on three different research datasets. The F1 score, recall, and precision of the designed model are 92.81%, 91.63%, and 94.46%, respectively. Nevertheless, the designed model lacked the capability to interpret contextual semantic variations and hidden emotions such as sarcasm and implicit opinions in textual reviews. [46] applied a technique to sentiment analysis that considered

user behavior through the usage of hotel reviews. Most of the research effort went into developing a recommendation system for the hotel sector, with the goal of allowing customers to browse the best hotels based on their personal preferences. By analyzing reviews according to accuracy and utilizing multiple text-based and sentiment classification algorithms, the recommendation system's performance was enhanced. However, the system mainly depended on traditional sentiment classification methods and failed to provide aspect-aware recommendations based on multiple review attributes.

2.3 Aspect-Based Sentiment Analysis (ABSA)

Additionally, [47] proposed a novel framework for rating prediction featuring tailored consumer preference recommendations, termed Aspect-Based Cross-Domain Personalized Recommendation (AsCDPR). The use of aspect-based features and their counterparts significantly improved the accuracy and F1-score metrics of sentiment analysis, with values of 1.6682% and 1.3657%. However, the framework faced challenges in accurately capturing implicit sentiments and contextual dependencies across multiple domains, which limited recommendation precision. [48] developed a method for ranking hotels using online review texts. The method involved the use of the AS-Capsules algorithm to estimate the sentiment strength of different aspects generated by web reviews. An interval cumulative distribution vector representing the best interval was utilized to formulate the optimization problem, whose solution yielded weights for the various aspects. The method was effective since it considered some particular features of reviews, hence customizing and making the recommendation more precise for people with different tastes. However, the model paid less attention to identifying new aspects within the reviews and their associated semantics.

2.4 Transformer-based / generative model

[49] proposed a GAN-based recommendation algorithm. For users with diverse needs, accuracy-based recommendations were employed to match products closely with user preferences. Experimental simulations were conducted on the MovieLens-100K dataset. The experimental results indicated that the proposed algorithm effectively improved accuracy. However, the framework primarily concentrated on recommendation accuracy and did not incorporate fine-grained ABSA from textual reviews. [50] proposed a hybrid ABSA framework to tackle issues in customer feedback analysis, which optimised RoBERTa for Sinhala as well as code-mixed data. The experimental results indicated that the model attained an accuracy of 92.3% and an F1-score of 0.89. Despite its strong performance, the framework was language-specific and might face difficulties in generalising across multilingual and cross-domain recommendation environments. [51] developed an approach that proposed hotels based on content and aspect-based review classification on sentiment analysis of reviews. The authors have begun by employing a weight-assigning methodology involving BERT. Many of the textual attributes that have been utilized here include word vectors and pre-trained word embeddings generated through BERT models. Thereafter, the researchers have segregated these review texts into several groups through fuzzy logic-based categorization methodology. The results revealed that the BERT model exhibited an accuracy rate of 92.36%. Nevertheless, the developed framework had a high computational complexity along with inefficient scalability.

3. Proposed NSAR-T Framework

The design process and implementation of the NSAR-T framework that builds the recommendation system by following the process of sequential development are described in the section below. The methodology creates a cohesive framework, whereby the system is able to incorporate the detailed user preferences during the process of making customized recommendations that can further be scaled to suit user requirements. The flow chart depicted in Figure 3 represents the entire operational process of the proposed NSAR-T framework.

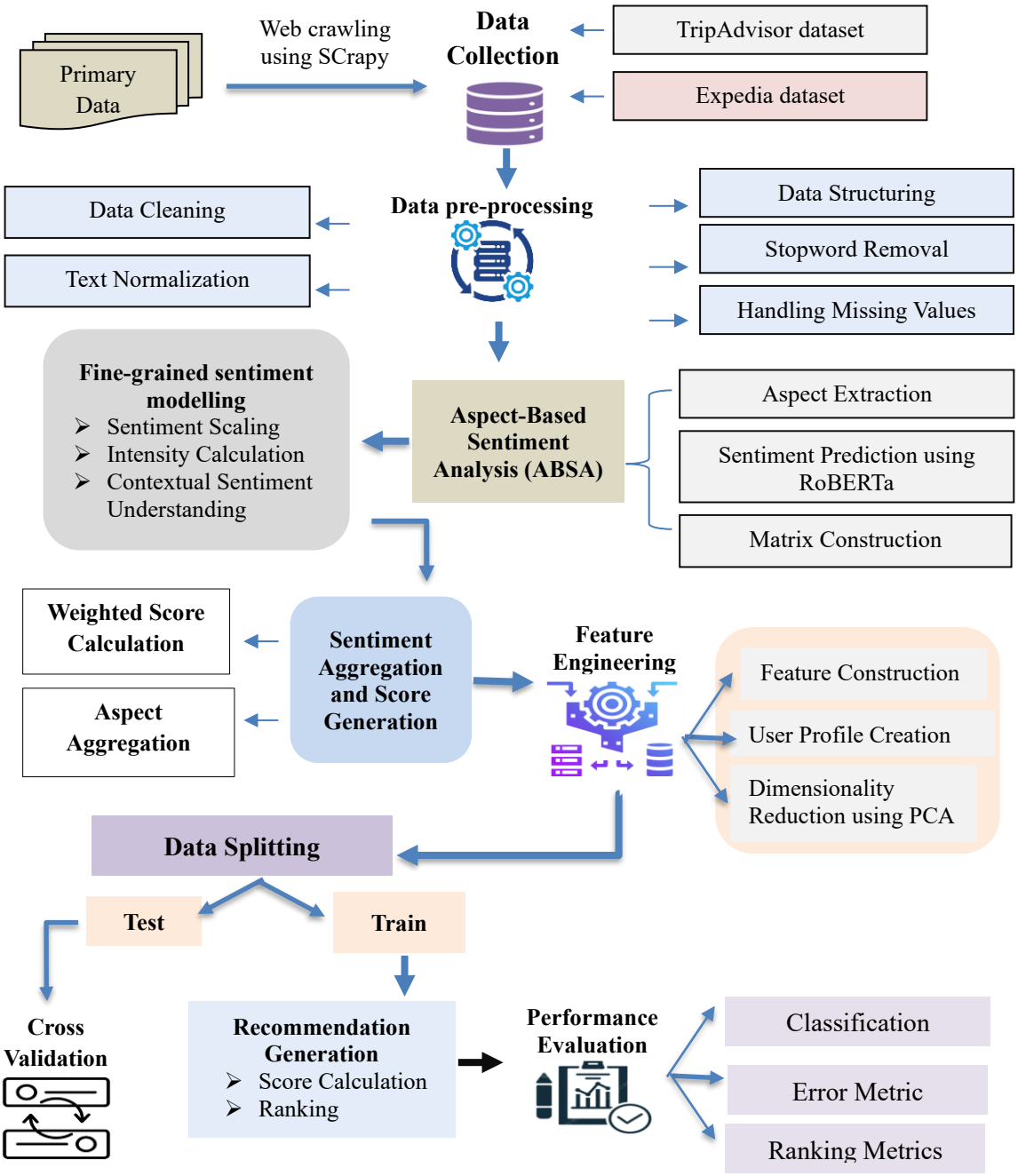

Figure 3: Proposed Framework

3.1 Dataset Description

This section shows the dataset used in this research:

3.1.1 Primary Dataset

This research employs the primary data set in the form of 200000 reviews created by users and collected using web scraping from several online travel websites. The research team applies the Scrapy and BeautifulSoup programs for collecting information from TripAdvisor, Expedia, and several other hotel booking websites. This data set is rich in unstructured text describing the user experiences and perceptions about the hotels' services, with examples such as "Excellent service and clean rooms", "Great location but disappointing food quality", and "Great experience waiting two hours for check-in". Purposeful sampling is employed in the selection of the data set, which ensures the preservation of one of the most fundamental criteria for the presence of positive, negative, neutral, sarcastic, and aspect-based reviews. The model requires this type of variety to enable it to understand different sentiments and ABSA.

3.1.2 Secondary Dataset

Two large-scale secondary datasets are incorporated to enhance the robustness, diversity, and scalability of the proposed framework.

- **Expedia Hotel Recommendation Dataset**

This study employs the Expedia Hotel Recommendations dataset, sourced from Kaggle [52]. The dataset consists of over 37.6 million training samples and 2.5 million testing samples that record real user searching behaviors and booking patterns for the hotels. User preference patterns that were included in the dataset consist of those who liked hotels with positive attributes, such as clean rooms, ideal location, and reasonable prices, and disliked hotels with negative service quality or negative user experience. Only about 2-3 million data points were used for the research due to computational constraints on processing the full data set. The dataset creates an imbalanced distribution that mirrors actual user patterns, thus enabling the development of strong recommendation systems that can handle large-scale operations.

- **TripAdvisor Hotel Reviews Dataset**

The TripAdvisor Hotel Reviews dataset, which was retrieved from the Kaggle dataset [53], contains 878,561 reviews from 4,333 hotels over 15 years. The review texts contain various forms of customer opinions and aspect-level sentiments, including statements such as “Great location but average food”, “The staff was friendly and supportive”, and “Rooms were clean, but the service was slow”. This dataset is a major source of research for ABSA tasks since it provides detailed sentiment information that helps researchers understand user preferences and aspect-specific opinions. Table 1 shows the dataset description used in this research.

Table 1: Datasets used

Dataset Name	Dataset Type	Dataset Size	Key Parameters / Features	Example Records / Attributes
Primary Web-Scraped Hotel Reviews Dataset	Primary Dataset	200,000 Reviews	Review Text, Rating, Timestamp, User Location, Stay Type, Price Range	“Excellent service and clean rooms”, Rating = 4.5, Location = USA
Expedia Hotel Recommendation Dataset	Secondary Dataset	37.6M Train / 2.5M Test Records	User ID, Hotel Cluster, Search Destination, Booking Status, Price, Device Type, Timestamp	User Location = 104, Hotel Cluster = 23, Booking = 1
TripAdvisor Hotel Reviews Dataset	Secondary Dataset	878,561 Reviews	Review Text, Rating, Hotel Name, Service Score, Cleanliness, Location Score	“Great location but average food”, Rating = 4

3.1.3 Data Preprocessing

Preprocessing provides an essential step that lays the foundation for ensuring data consistency and data reliability. Real-life data collected in the form of textual data is subject to various forms of noise as it includes various ways of writing, various source components, and several formats of data. In order to make the data consistent, preprocessing cleans up the data by discarding unnecessary components like Hypertext Markup Language (HTML) tags and special characters. The cleaned review is represented as:

$$r'_{ij} = \text{Clean}(r_{ij}) \tag{1}$$

where r'_{ij} denotes the processed textual input. Transformation reduces lexical diversity but increases the semantic comprehension ability; this is a key prerequisite for future language modelling. The dataset maintains its original configuration since statistical imputation is employed to deal with the lack of interaction value in the data. This provides a very crucial basis for making the learning process efficient and minimizing biases.

3.2 Aspect Extraction and Sentiment Classification

To capture the multi-dimensional nature of user preferences, the proposed NSAR-T framework employs ABSA integrated with RoBERTa-based transformer modelling for fine-grained opinion understanding.

3.2.1 ABSA

ABSA breaks down customer reviews into multiple opinion-based components, which allows users to understand their sentiments about particular reviews instead of using the traditional method that assigns one sentiment to an entire review [54]. The framework detects user sentiments about specific review elements, including service quality and cleanliness, pricing, and location and amenities, which helps users understand their preferences in a detailed and understandable way. The extracted aspects from each review are represented as:

$$A_{ij} = \{a_1, a_2, a_3, \dots, a_k\} \tag{2}$$

where A_{ij} denotes the set of extracted aspects from the review r_{ij}, and $a_k$represents an individual aspect identified within the review text.

3.3.2 RoBERTa

The framework uses RoBERTa as its main transformer-based model for sentiment analysis to determine the sentiment polarity of each aspect, which was extracted after the system completed aspect extraction. RoBERTa is an advanced transformer architecture derived from BERT that improves language representation learning through optimized pre-training strategies and enhanced contextual embedding generation [55]. The model uses a bidirectional self-attention system, which processes text by attending to all surrounding words at the same time to extract long-distance semantic links, contextual connections, and hidden language patterns from review text [56]. The advanced capabilities of RoBERTa enable it to process complex sentence structures and track changes in emotional context, and detect sarcasm and subtle sentiment differences, which standard machine learning methods and basic neural networks cannot manage.

The sentiment polarity for each extracted aspect is computed as:

$$s_{ijk} = f_\theta(r'_{ij}, a_k) \tag{3}$$

where s_{ijk} denotes the sentiment polarity score corresponding to the aspect a_k, r'_{ij} represents the preprocessed review text, and f_θ denotes the RoBERTa transformer model. With this combination of the ABSA-RoBERTa approach, the model converts unstructured texts into structured aspect-based sentiment representation. This allows sentiment analysis to be done accurately, which facilitates personalized recommendation generation.

3.4 Fine-Grained Sentiment Modelling

The framework develops a new approach to sentiment analysis that provides detailed sentiment measurement through its fine-grained system. The model uses continuous sentiment measurement, which enables it to identify different levels of user satisfaction instead of using specific sentiment categories that include positive, negative, and neutral. The system uses this capability to identify different sentiment levels, which include mild dissatisfaction and strong dissatisfaction that typical models fail to detect. The system uses continuous sentiment values to create seven semantic levels, which enhance its interpretability according to Table 2.

Table 2: Classification of sentiment level

Sentiment Score Range	Sentiment Level
+0.75 to +1.00	Strongly Positive
+0.50 to +0.75	Positive
+0.25 to +0.50	Nearly Positive
-0.25 to +0.25	Neutral
-0.50 to -0.25	Nearly Negative
-0.75 to -0.50	Negative
-1.00 to -0.75	Strongly Negative

In the designed approach, the input of the fine-grained sentiment classification component includes the processed reviews and the review aspects. In contrast, the output of the sentiment analysis component includes the sentiment scores of the seven different sentiments that are assigned to the review aspects. With the help of this type of input and output process, the designed recommendation framework is able to provide classifications based

on sentiment intensity levels and also convert the opinion text into numerical values. The detailed sentiment outputs produced during the sentiment analysis phase serve as the primary input for the recommendation generation phase. Based on the degree of sentiments, the items classified under the positive sentiment category, such as strong positive, positive, and near-positive, would be recommended to the consumers as opposed to the neutral and negative sentiment categories.

3.5 Sentiment Aggregation and Score Generation

After the sentiment is classified at an aspect level, the sentiment aggregation and score calculation processes follow, enabling a shift from aspect-based sentiments towards a unified sentiment that serves as a recommendation for the user. Since one review contains many aspects of varied sentiment intensities, the sentiment analysis framework adds up the sentiment score for each sentiment based on its respective importance in the evaluation process. This sentiment score calculation process involves assigning weight values differently to important and less important sentiments. The sentiment aggregation is computed as:

$$S_{ij} = \sum_{k=1}^{K} w_k \cdot s_{ijk} \tag{4}$$

where S_{ij} represents the final aggregated sentiment score for the user u_i toward item v_j, s_{ijk} denotes the sentiment score associated with the aspect a_k, and w_k represents the weight assigned to each aspect according to its importance. The output generated from sentiment analysis is achieved from the sentiment scores produced in the process, which serve as the main input for the recommendations phase. The structure of the sentiment score enables qualitative text data into a form that can be expressed numerically and ranked.

3.6 Feature Engineering

The feature engineering step designs a structured sentiment-based feature representation by combining the extracted sentiment scores and intensities along with the aspect importance weight generated from the fine-grained sentiment analysis step. The use of such designed features allows the framework to capture the complex patterns of user preferences on multiple aspects, while at the same time maintaining semantic connections among users' perceptions regarding item features. In order to enhance the computational efficiency of the model, PCA is applied to the created feature space.

$$U_i' = PCA(U_i) \tag{5}$$

where U_i' represents the reduced feature representation. PCA helps to project the highly dimensional feature space into a reduced dimensional space, while retaining the maximum variability of information from the dataset [57]. The process of dimensionality reduction makes it possible to reduce noise, increase efficiency, and enable scalability in recommendations for big datasets.

3.7 Data Splitting and Validation

The dataset is split in a ratio of 80:20; thus, two datasets are developed that would be used for training and testing purposes, since the training data of 80% needs to be tested by the 20% dataset. Model Evaluation involves the use of novel data that has never been seen before to examine the capability of the model to extrapolate from the training data. Model Parameters are learned using training data, while testing uses the test data. Training of the model is conducted using the k-fold CV technique, which enhances model performance and minimizes the risk of overfitting. In the training of the model, k parts are created from the training data, and model training utilizes k-1 parts, leaving one as validation data. This process is conducted k times and ensures that all the k parts are used as validation data at some point during training.

3.8 Recommendation Generation

The final recommendation stage utilizes the computed sentiment outputs from fine-grained sentiment modelling to generate personalized recommendations based on user opinions and aspect-level sentiment preferences. The recommendation score is computed as:

$$Score(u_i, v_j) = \sum_{k=1}^{K} \lambda_k S_{ijk} \tag{6}$$

where S_{ijk} represents the sentiment score associated with aspect k for the user u_i toward item v_j, and λ_k denotes the weighting coefficient representing the relative importance of each aspect in the recommendation process. The recommendation score model takes into account the ranking of all the items on the basis of sentiment scores that relate to the corresponding aspect of each user. The top N items with the highest rankings are selected as recommendation suggestions after calculating the sentiment scores. Through this ranking process based on sentiment, the generation of recommendation suggestions becomes possible because of a clear understanding of user sentiments and preferences related to different aspects.

3.9 Performance Evaluation

The proposed framework of NSAR-T is tested by using different measures, including its capacity for predicting outcomes, ordering items, and making recommendations. The framework employs the Classification-based measures that comprise accuracy, precision, recall, and F1-score. Prediction accuracy is evaluated using RMSE:

$$RMSE = \sqrt{\frac{1}{N}\sum(y_i - \hat{y}_i)^2} \tag{7}$$

where y_i is the actual interaction value, $\hat{y}_i$ is the predicted value, and N is the total number of predictions. RMSE indicates how close the predicted ratings are to the actual values.

The ranking assessment evaluation methodology incorporates both NDCG and MAP as its evaluation parameters. NDCG is measured as a metric of how well the recommended items are ranked, assigning more significance to recommendations at the top positions. MAP is calculated as the average precision of all users. The quality of the recommendations is evaluated based on three different parameters: Diversity, Coverage, and Novelty. Diversity is estimated based on similarity between the recommended items, while Coverage stands for the number of recommended items derived from the dataset.

3.10 Proposed Algorithm

This section shows the pseudo-code based on the proposed methodology.

Input:

Dataset D = {(u_i, v_j, r_ij)}

Output:

Top-N recommended items for each user

Begin

// **Step 1: Data Collection**

Load dataset D

// **Step 2: Data Preprocessing**

For each review r_ij in D do

Clean text → r_ij'

Apply tokenization and lemmatization

Remove stopwords and noise

End For

// **Step 3: Aspect extraction and Sentiment Classification**

For each cleaned review r_ij' do

Extract aspects:

A_ij = {a1, a2, ..., ak}

For each aspect a_k in A_ij do

Compute sentiment score:

s_ijk = f_θ(r_ij', a_k)

End For

End For

// **Step 4: Fine-Grained Sentiment Modelling**

For each aspect, sentiment score s_ijk do

Normalize sentiment into a seven-level scale:

{Strongly Positive, Positive, Nearly Positive,

Neutral, Nearly Negative, Negative,

Strongly Negative}

End For

// **Step 5: Sentiment Aggregation and Score Generation**

For each user-item pair, do

Compute aggregated sentiment score:

$S_{ij} = \Sigma (w_k * s_{ijk})$

End For

// **Step 6: Feature Engineering**

Construct sentiment-aware feature vector:

$U_i = \{S_{ij}, A_{ij}, w_k\}$

Apply dimensionality reduction:

$U_i' = PCA(U_i)$

// **Step 7: Data Splitting and Validation**

Split dataset D into:

$D_{train} = 80\%$

$D_{test} = 20\%$

Apply k-fold cross-validation

// **Step 8: Recommendation Generation**

For each user u_i and item v_j do

Compute recommendation score:

$Score(u_i,v_j) = \Sigma (\lambda_k * S_{ijk})$

End For

Rank items based on Score

Select Top-N items

// **Step 9: Performance Evaluation**

Compute:

Accuracy, Precision, Recall, F1-score

RMSE

NDCG, MAP

End

4 Results and Discussion

This section illustrates the results achieved during experiments with the suggested approach on various datasets as well as sentiment-aware recommendation measures. Table 3 depicts hyperparameters and configuration parameters that were set to create the sentiment-aware personalized recommendation generation system. As a model to perform contextual sentiment analysis, RoBERTa transformer is used in the framework, together with a seven-level sentiment modeling technique to detect nuances in sentiment. The algorithm of PCA is applied to reduce dimensionality to 13 components. The 80/20 train/test split and five-fold CV are performed so that model validation can be robust enough. As for recommendations, a Random Forest classifier is employed with 120 decision trees, an eight-maximum depth, and tuned split parameters in order to increase recommendation accuracy and improve generalization capability.

Table 3: Hyperparameters used

Module	Important Hyperparameters	Value
Sentiment Model	RoBERTa Model	cardiffnlp/twitter-roberta-base-sentiment-latest
Aspect Analysis	Sentiment Classes	Positive, Neutral, Negative
Fine-Grained Sentiment	Sentiment Levels	7 Levels
Feature Engineering	PCA Components	13
Data Splitting	Train-Test Ratio	80:20
Cross Validation	K-Fold CV	5-Fold
Recommendation Model	Classifier	Random Forest
	Number of Trees (n_estimators)	120
	Maximum Depth (max_depth)	8
	Minimum Samples Split	8
	Minimum Samples Leaf	4
	Feature Selection	sqrt
Randomization	Random State	42

Figure 4 demonstrates the process of collecting the dataset as well as the size of each of the datasets applied to the proposed NSAR-T framework. The Primary dataset consists of 200,000 records, while the TripAdvisor Dataset consists of 878,561 records. On the other hand, the Expedia Dataset consists of 37,670,293 records. As can be seen from the figure below, the collected data is on a massive scale.

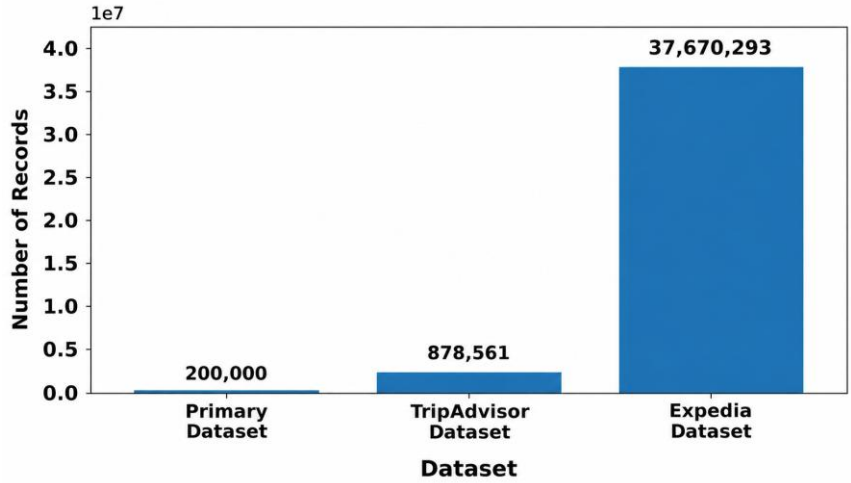

Figure 4: Data Collection

The results of data cleaning are shown in Figure 5 below. This figure illustrates the change in the record numbers in the Primary, TripAdvisor, and Expedia Datasets before and after data cleaning. In particular, the Primary Dataset lost 1,095 records while reducing from 200,000 to 198,905 records; the TripAdvisor Datasets were reduced by 5,974 records from 878,561 to 872,587; and the Expedia Dataset had a decrease of 37,470,295 records.

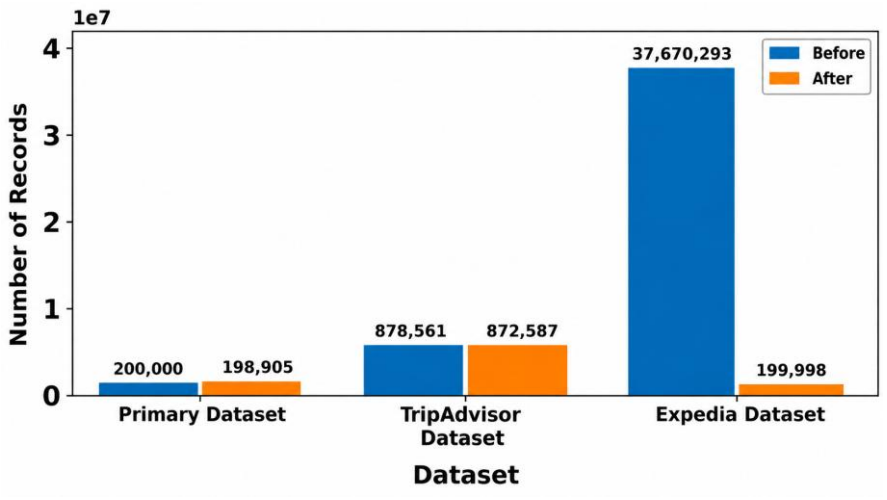

Figure 5: Data Cleaning

Figure 6 represents the mean value of results that were generated by implementing various preprocessing strategies within the framework developed. It is clear from the graph that Data Cleaning generated the highest mean value at 423,830, while Handling Missing Values came second at 22,177. On the other hand, Data Structuring, Stopwords Removal, and Text Normalization had mean values of 18, 46, and 51, respectively.

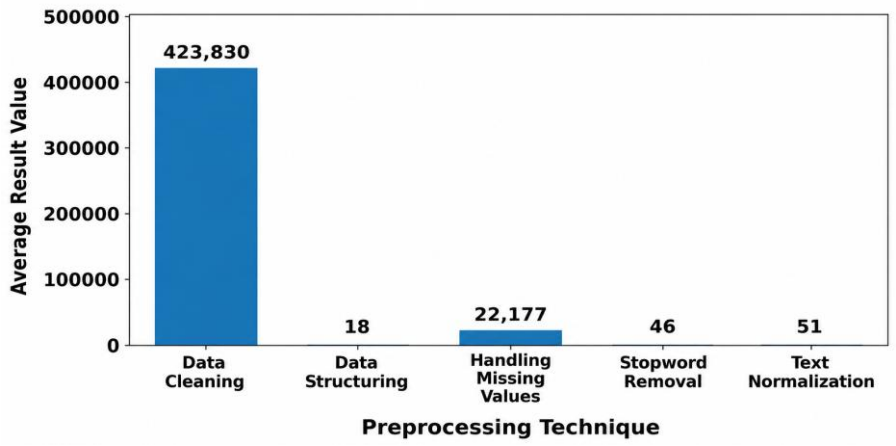

Figure 6: Average result values after preprocessing

Figure 7 presents the extracted aspect distribution for two datasets, because ABSA was done only on the textual review datasets, such as the Primary Dataset and the TripAdvisor Dataset. The Expedia dataset was not included here, as it mainly contains structured recommendations and booking-related data rather than detailed textual reviews needed for extracting opinion aspects and sentiments. Figure 7 shows that location and staff aspects reached the highest counts in both datasets. In the Primary Dataset, location recorded 5,885 instances and staff 5,467 instances, while in the TripAdvisor Dataset, staff reached 6,214 instances and location 5,391 instances, which suggests that users discussed hotel location quality and staff performance most often in their reviews.

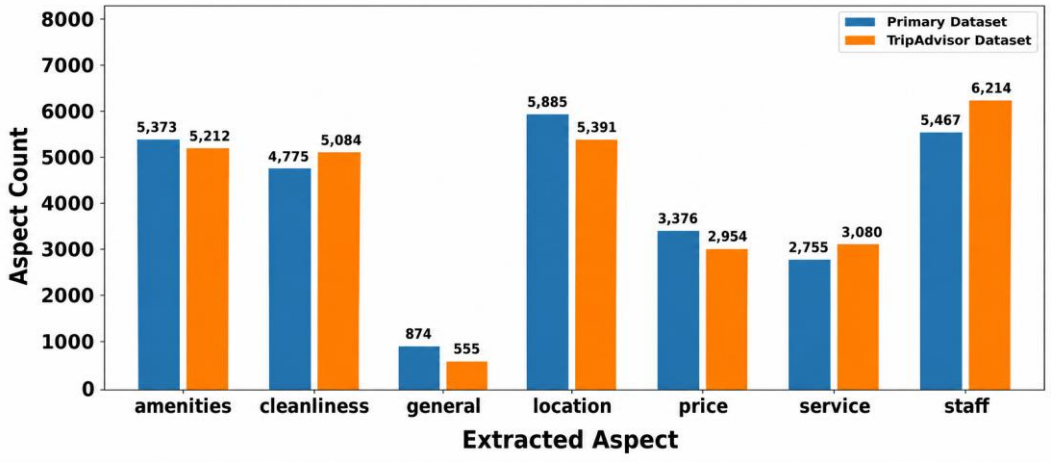

Figure 7: Extracted Aspect Distribution

Figure 8 shows the average sentiment scores for various hotel aspects across the Primary Dataset and the TripAdvisor Dataset. The graph suggests that cleanliness, amenities, location, and staff got comparatively higher sentiment scores in both datasets, so people seem to give more favorable impressions about those hotel attributes. In the Primary Dataset, cleanliness came in with the top average sentiment score of 0.683988, and in the TripAdvisor Dataset, cleanliness is at 0.665178. Overall, this indicates strong customer satisfaction with hotel cleanliness and the related services.

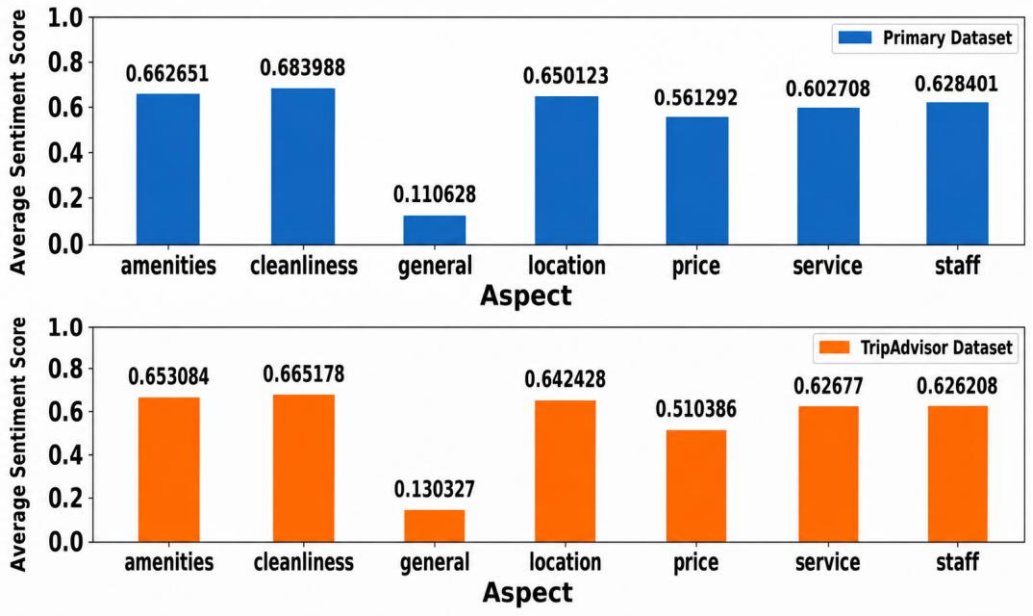

Figure 8: Average Sentiment Score

Figure 9 shows the sentiment distribution produced by a RoBERTa-based sentiment classification model for the Primary Dataset and the TripAdvisor Dataset. The plot basically suggests that most reviews were labeled as positive sentiments in both sets. For the Primary Dataset, there were 21,947 positive reviews, 4,682 neutral reviews, and 1,876 negative reviews. Meanwhile, the TripAdvisor Dataset reported 22,051 positive reviews, 4,224 neutral ones, and 2,215 negative reviews, so overall it seems like customers had a favorable opinion about hotel services and experiences.

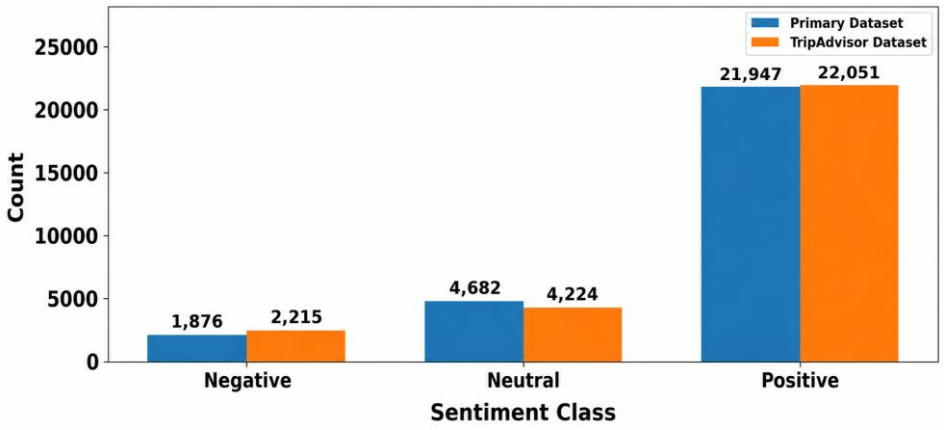

Figure 9: RoBERTa Sentiment Distribution

Figure 10 shows the Primary Aspect Sentiment Matrix, which was generated with the RoBERTa sentiment classification model. Basically, this matrix illustrates how negative, neutral, and positive sentiments are distributed across various extracted hotel aspects. The outcome indicates that positive sentiments dominate most aspects, and especially location with 4,657 positive reviews, amenities with 4,293 positive reviews, and staff with 4,292 positive reviews. On the other hand, the general aspect comes out differently, with comparatively lower positive sentiment counts but higher neutral sentiment values. This usually suggests customers had more mixed feelings about the overall hotel experience.

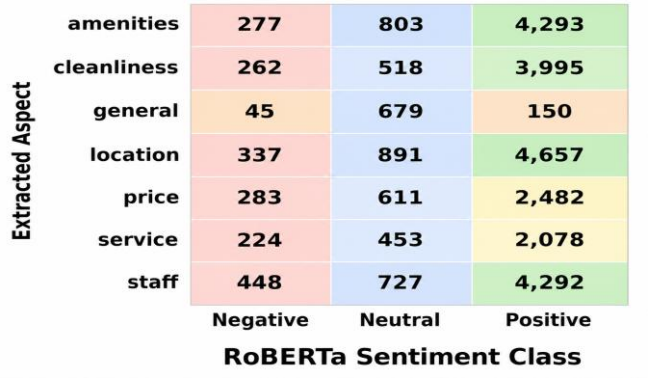

Figure 10: Primary Aspect-Sentiment Matrix

Figure 11 shows the TripAdvisor Aspect Sentiment Matrix, which was put together using a RoBERTa sentiment classification model. In general, the outcomes suggest that positive sentiments take the lead across most aspects, especially staff, with 4,880 positive reviews, location, with 4,231 positive reviews, and amenities, with 4,176 positive reviews. Meanwhile, the overall aspect ends up with lower positive sentiment figures, and this kind of thing usually points to more blended views from customers about the overall hotel experience.

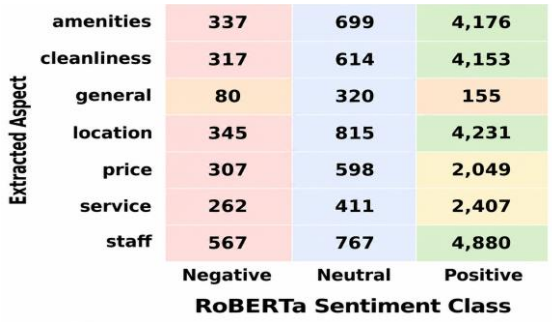

Figure 11: TripAdvisor Aspect-Sentiment Matrix

Figure 12 shows the sentiment intensity distribution that came out during the fine-grained sentiment modelling process for both the Primary Dataset and the TripAdvisor Dataset. As can be seen from the graph, the occurrence rates of intense sentiment are the maximum in both data sets, being 14,796 for the Primary data set and 15,585 for the TripAdvisor data set. However, medium intensity and neutral sentiments also have some values,

whereas low intensity seems to be less in number. This suggests that most user reviews expressed strong emotional opinions toward hotel experiences and services.

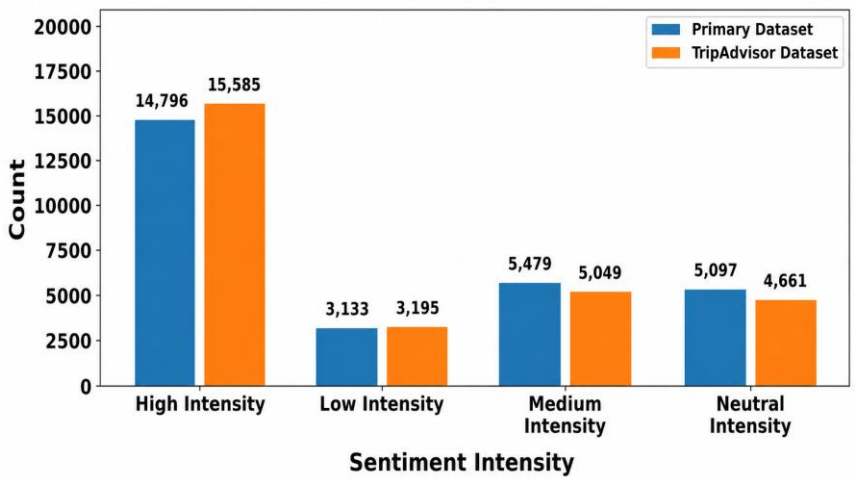

Figure 12: Sentiment Intensity Distribution

Figure 13 shows the distribution of fine-grained sentiment levels produced by the proposed sentiment modeling framework, for both the Primary Dataset and the TripAdvisor Dataset. The graph indicates that Strongly Positive sentiments got the highest counts in both datasets, 14,505 instances in the Primary Dataset and 15,157 instances in the TripAdvisor Dataset. Positive and Neutral sentiment levels also show rather strong values, and then Strongly Negative sentiments are the ones with the lowest counts, which is suggestive that most user reviews basically reflect very positive views about hotel services and the whole experience.

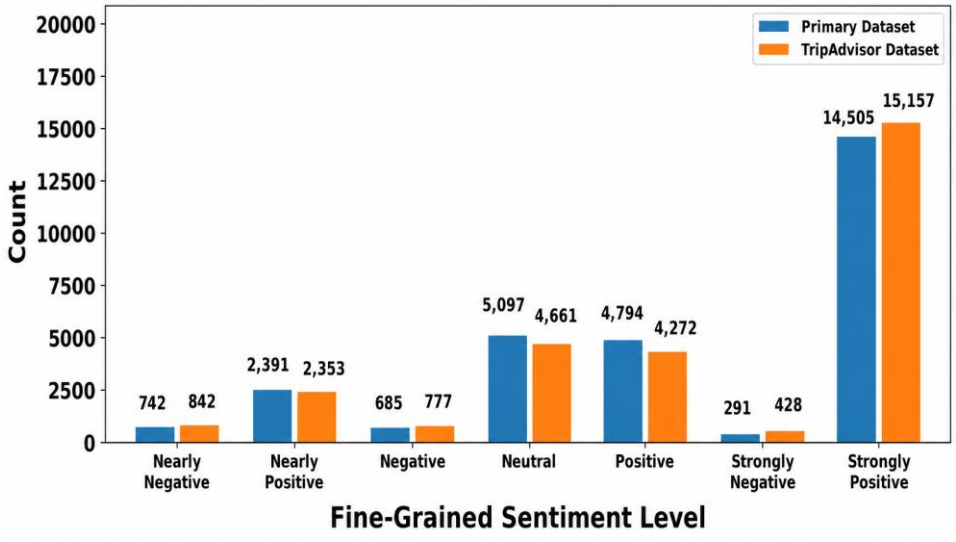

Figure 13: Sentiment Level Distribution

Figure 14 presents the contextual sentiment matrix created in the fine-grained sentiment modeling phase. According to the proposed approach, sentiment scores generated by RoBERTa have been somehow grouped into seven smaller sentiment classes, capturing those small emotional variations for each aspect of the hotels, rather than having only a broad perspective. As can be seen in the matrix presented, the number of Strongly Positive sentiments seems to be the highest overall, with particularly high numbers for the staff (3,507), location (2,930), and amenities (2,832) aspects. Meanwhile, Strongly Negative sentiments sit at the very bottom, with the smallest counts, suggesting that most TripAdvisor reviews leaned toward really good customer experiences and overall satisfaction levels.

Figure 14: TripAdvisor contextual sentiment matrix

The matrix for the primary contextual sentiment analysis is provided in Figure 15. It can be observed from the Primary context matrix that the Strongly Positive sentiment occurs mostly in all parameters. Strongly positive sentiment occurred at the highest frequency in the following parameters: location 3,120, staff 2,923, and amenities 2,912. However, the occurrence of neutral sentiment and positive sentiment was also high. On the other hand, the occurrence of strongly negative sentiment was low.

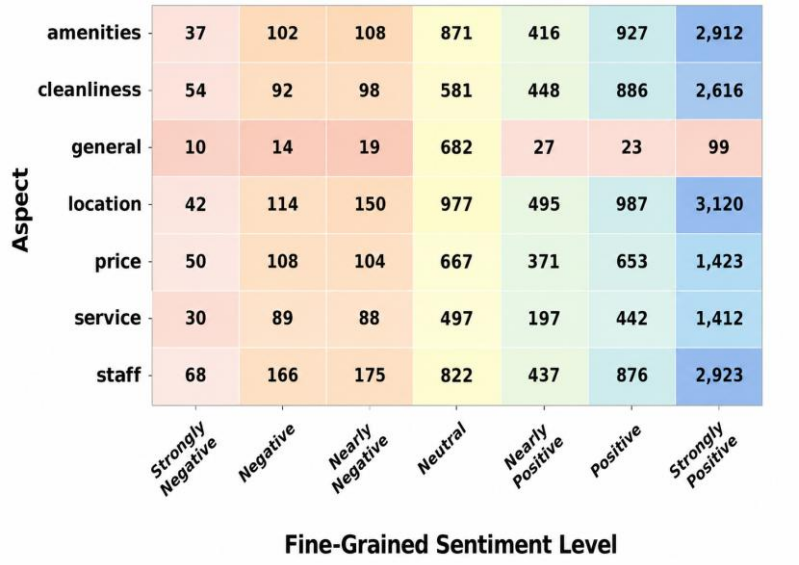

Figure 15: Primary contextual sentiment matrix

Figure 16 shows the aspect-wise sentiment strength that was produced during the fine-grained sentiment aggregation stage. In line with the methodology, the framework computes average absolute sentiment strength values for every extracted aspect, so it can gauge the perceived intensity of user opinions across the various hotel attributes. Looking at the graph, cleanliness, staff, and amenities end up with noticeably higher sentiment strength values in both datasets. For the Primary Dataset, cleanliness reaches the strongest value at 0.668. Meanwhile, for the TripAdvisor Dataset, staff posts the top score of 0.670. That basically points to stronger positive reactions from customers regarding hotel cleanliness, staff conduct, and overall service quality.

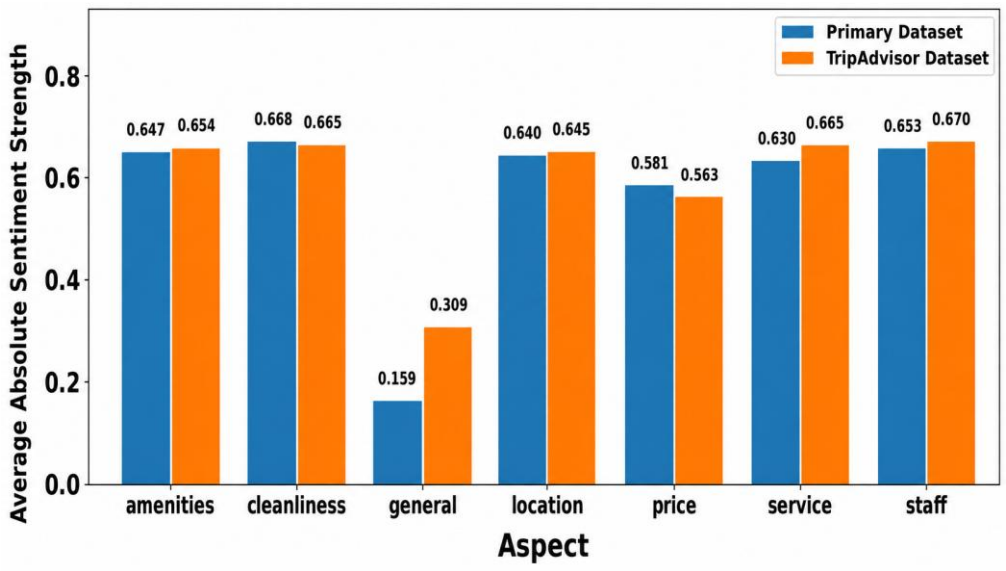

Figure 16: Aspect-wise Sentiment Strength

Figure 17 depicts the comparison of enhanced sentiment components and the Expedia score component for recommendation generation assistance. The Sentiment Component refers to the combined values of sentiments that are attained from the primary and TripAdvisor data set by means of fine-grained sentiment analysis and achieves the highest average score of 0.613. The Expedia Component refers to the values attained from structured data of Expedia recommendations and acquires an average score of 0.277. Then the Enhanced Recommendation Score, that is the one that merges all three datasets, reaches a more moderate average score of 0.458, and it indicates the approach works well by pulling together both the textual sentiment information and the structured recommendation data, so the overall personalized recommendation performance improves.

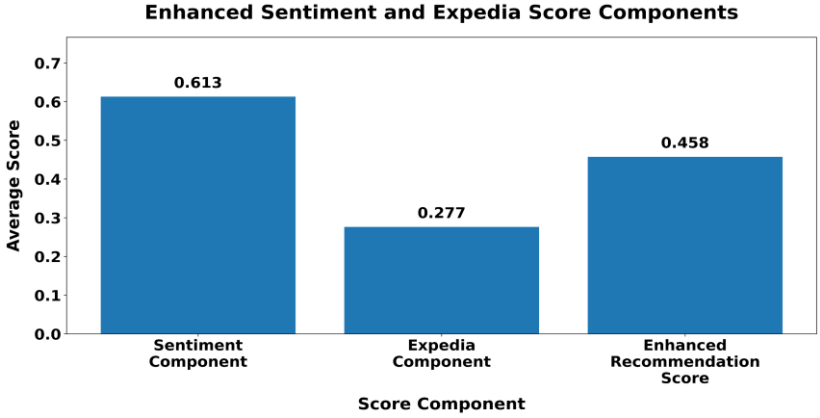

Figure 17: Average Sentiment Score

Figure 18 shows the distribution of the enhanced sentiment scores produced in the process of sentiment aggregation and score generation. This histogram shows the distribution of the enhanced sentiment scores that have been computed after consolidating the sentiment scores extracted from the Primary Dataset, TripAdvisor Dataset, and structured Expedia recommendations to form one sentiment-aware score. It is apparent from the graph that the distribution peaks at the midrange between 0.3 and 0.6. Most sentiment scores lie in this range, implying that most of the reviews and recommendations were favorable.

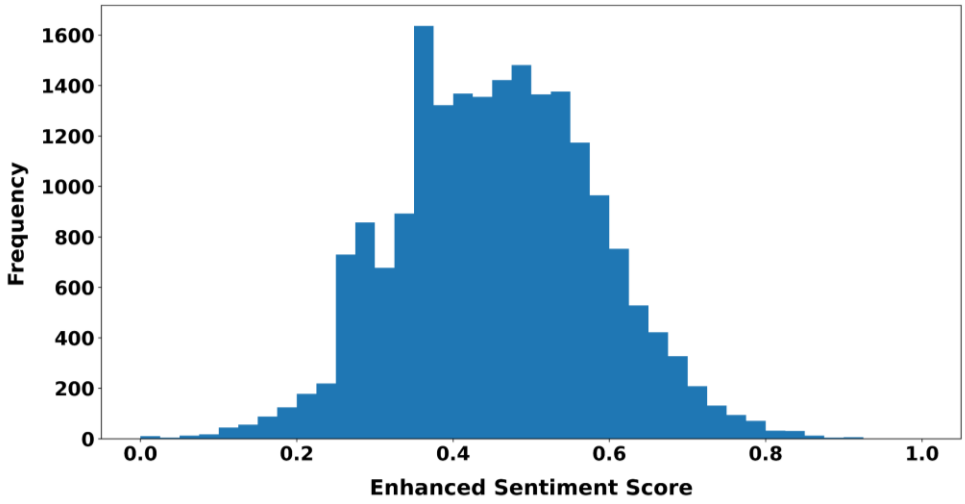

Figure 18: Enhanced sentiment score distribution

Figure 19 shows the cumulative explained variance calculated at the feature engineering phase using PCA. It depicts the contribution of each principal component in preserving the overall variance of the dataset. The cumulative explained variance slowly rises from 0.536 in Principle Component (PC1) to 0.979 in PC8, which means that 97.9% of all dataset variance is preserved using the selected principal components. The findings show the efficiency of using PCA for reducing the dimensions while preserving the most relevant sentiment-based feature information required for recommendation creation.

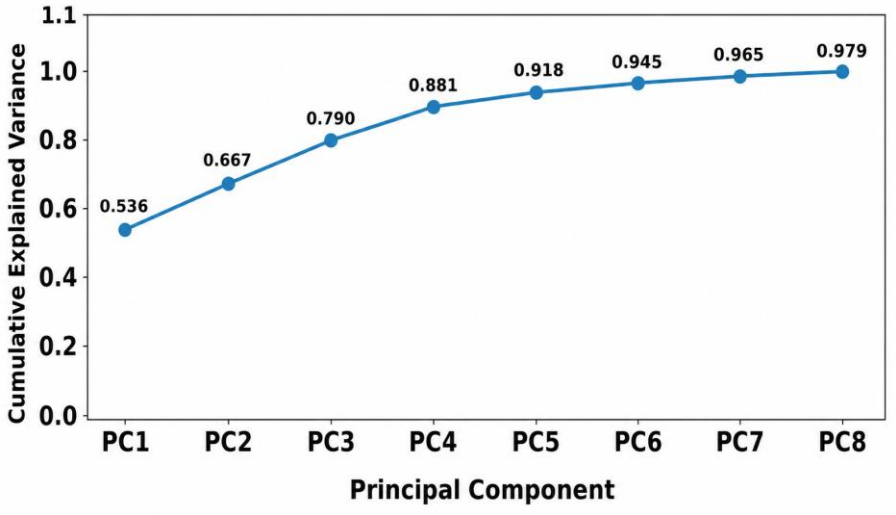

Figure 19: Cumulative Explained Variance

Figure 20 shows the explained variance ratio, which is derived in the feature engineering process. This figure highlights the variance value that is contributed by each principal component for maintaining the variance of the dataset. PC1 explained variance ratio was at its peak with 0.536, while PC2 and PC3 were second and third with ratios of 0.130 and 0.123, respectively. Other principal components explained less variance values, meaning that most of the features with sentiment awareness were captured in the first few principal components.

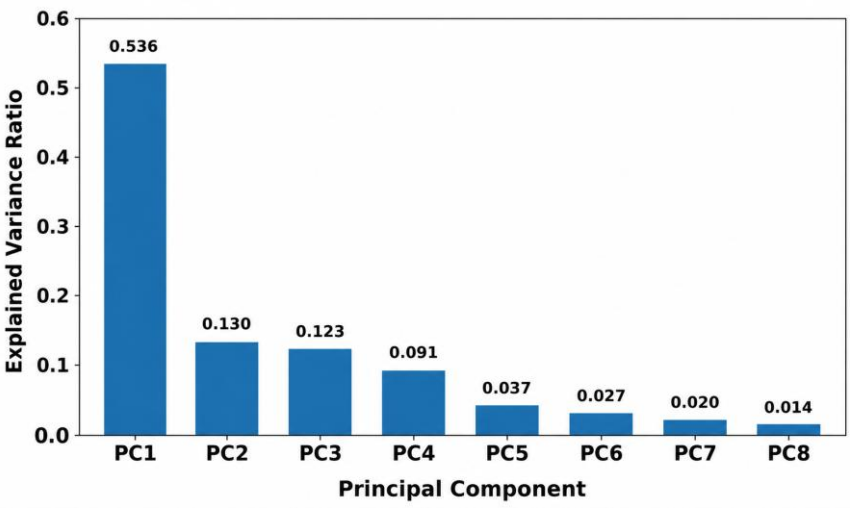

Figure 20: Explained Variance Ratio

The graph in Figure 21 shows the average values of the engineered features from the process of feature engineering. The figure depicts the importance of various engineered features employed to make recommendations. For instance, the feature recommendation_numeric had an average value of 3, while aspect_density_score had an average of 2. Other features, such as Positive_aspect_ratio, had an average value of 0.714, whereas sentiment_strength had an average value of 0.297. On the other hand, negative_aspect_ratio and rating_sentiment_interaction were relatively low.

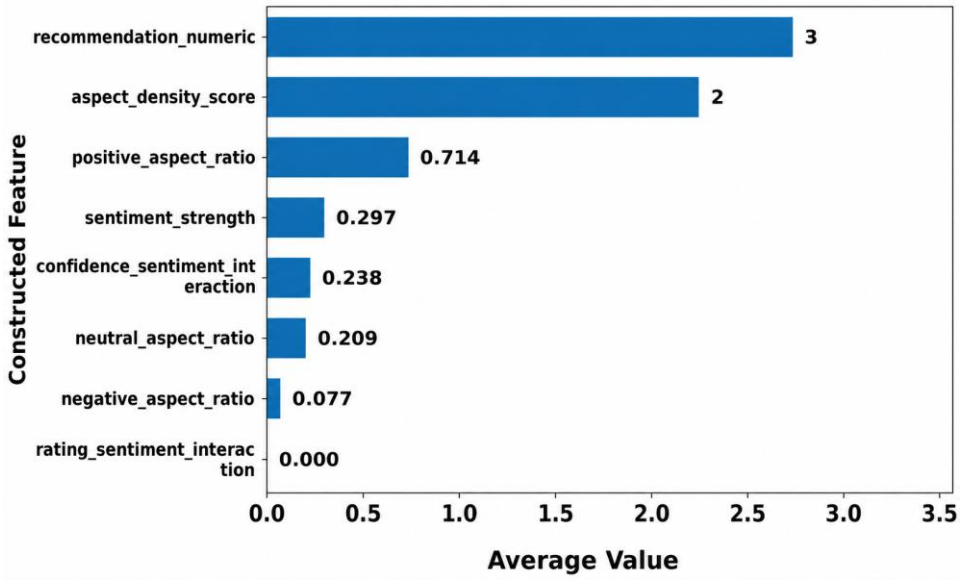

Figure 21: Average Engineered Feature Values

Figure 22 shows the dataset splitting, which was done in the validation stage. There is a graph that basically shows how the processed dataset gets divided into training and testing parts, for evaluation, and also for performance analysis. In the training part, there are 16,000 records, and for testing, there are 4,000 records. So, it works out to an 80:20 split ratio. This way, there is enough data available for learning, but at the same time, there is also an independent subset to check recommendation accuracy and generalization performance.

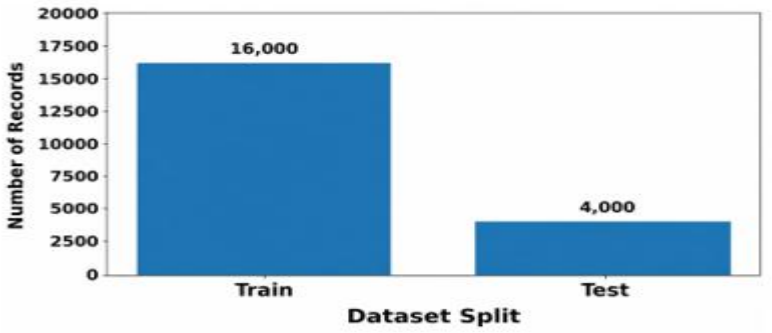

Figure 22: Data splitting

Figure 23 shows the distribution of the recommendation classes following the division of the dataset into training and testing sets. Figure 23 indicates the distribution of recommendation classes from 0 to 5 in terms of the train and test data sets. Classes 2 and 3 showed the highest number of occurrences with 6,942 and 5,512 observations in the train data set and 1,735 and 1,378 observations in the test data set, respectively. Conversely, classes 0 and 1 registered relatively fewer occurrence, indicating that the majority of recommendations belonged to moderate and positive recommendation categories.

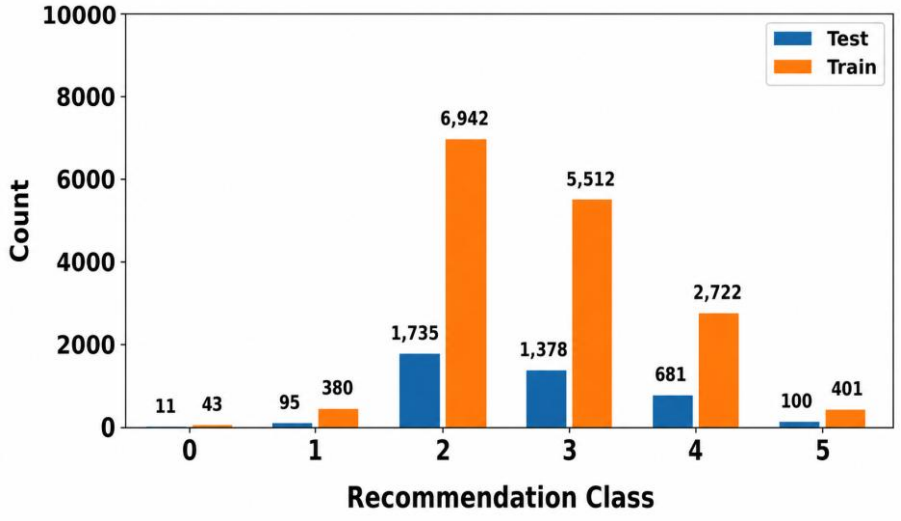

Figure 23: Recommendation class distribution

Figure 24 presents the CV train class distribution across five folds used during the model validation process. The graph illustrates the distribution of recommendation classes from 0 to 5 within each training fold. The second and third recommendation classes obtained the maximum number of recommendations, with about 5,553-5,554 and 4,409-4,410 observations, respectively. On the other hand, the rest of the recommendation classes have smaller numbers but relatively uniform counts. This implies that the use of cross-validation helped in achieving consistent class counts during the training process.

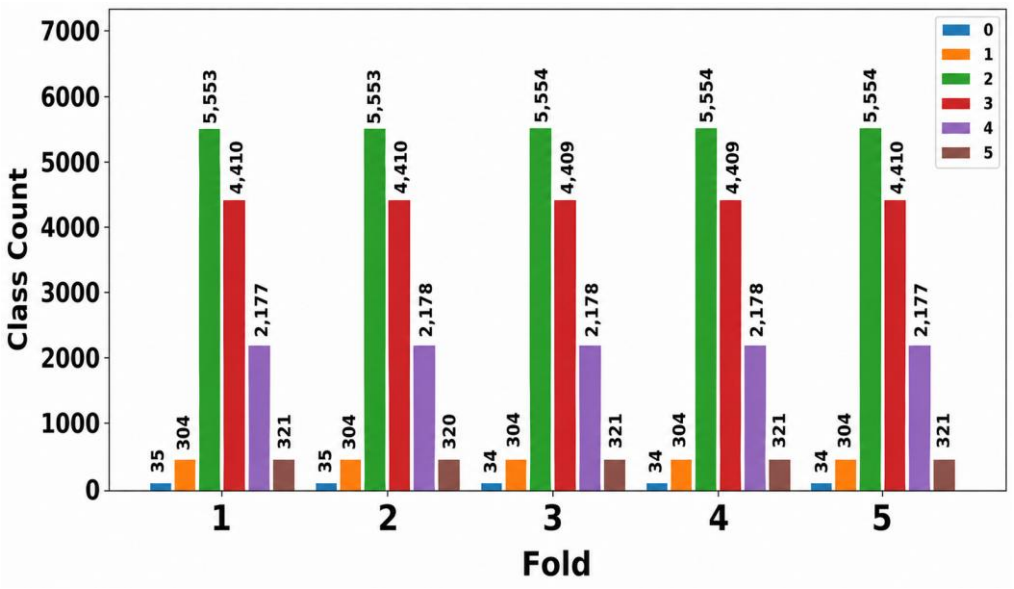

Figure 24: CV train class distribution across folds

Figure 25 shows the CV validation sample classification distributions for five folds of model evaluation. The recommendation class distributions for folds range from 0 to 5. Across all folds, recommendation classes 2 & 3 had the highest counts with about 1388 - 1389 records of class 2 and 1102 - 1103 records of class 3. All other classes were approximately evenly distributed among each fold with stable distributions and consistent validation accuracy across all folds of cross-validation, indicating a balanced partitioning of data for performance evaluation.

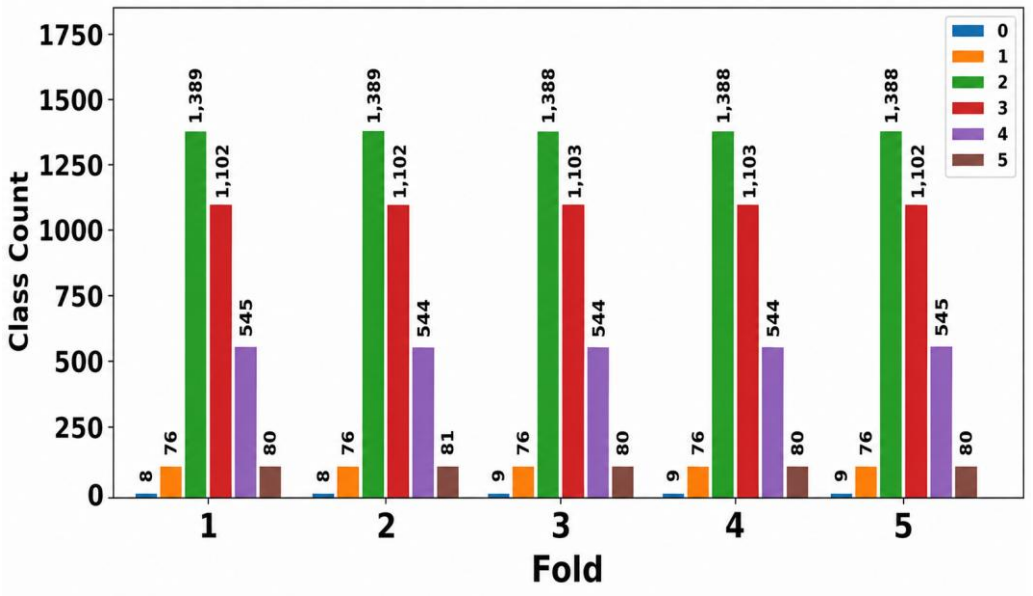

Figure 25: CV validation class distribution across folds

The Top-N Recommendation Scores generated in the recommendation ranking process can be seen in Figure 26. This figure shows the ranking of the top 10 recommended items by their final recommendation score. For example, the highest-ranking recommendations received the highest possible score of 1.000 for Rank 1, Rank 2, and Rank 3; all other recommendations received a score above 0.995. Because the recommendation scores are consistently high, this demonstrates that the framework can generate very relevant and personalized recommendations, based on summation sentiment-aware features.

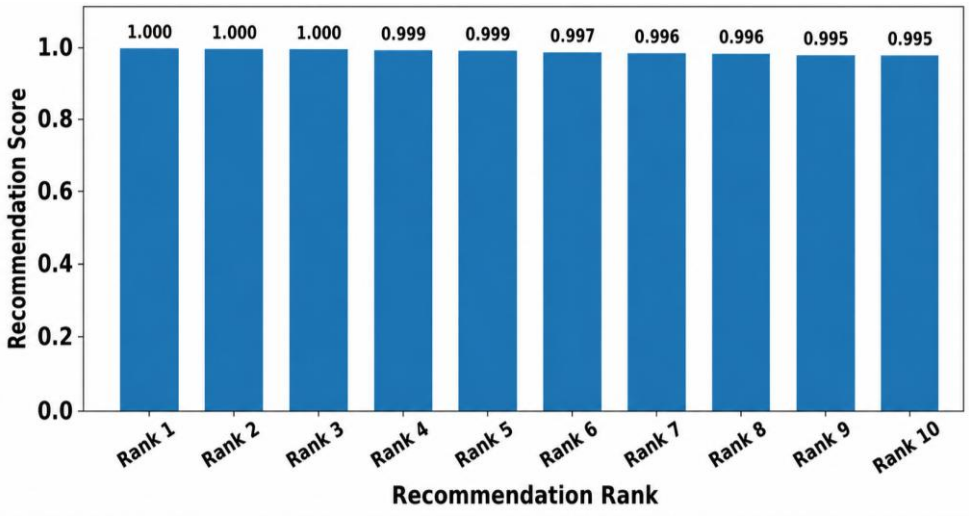

Figure 26: Top-N Recommendation Score

Table 4 shows the 5-fold cross-validation performance results of the proposed NSAR-T framework across different validation folds, like pretty much all of them. However, in all cases, the model maintained high accuracy, precision, recall, and F1-score, thereby making the classification stable and reliable. The stability of the ROC-AUC was also evident since its value remained constant close to 1.0, thereby proving the strength and robustness of the predictive capabilities of the model. Low RMSE and MAE further prove low levels of errors in the predictions.

Table 4: 5-Fold Cross-Validation Classification Metrics

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
Accuracy	0.998438	0.999688	0.999375	0.997188	0.998438
Precision	0.998841	1.000000	1.000000	0.997683	0.997688
Recall	0.998263	0.999421	0.998842	0.997105	0.999421
F1-Score	0.998552	0.999710	0.999421	0.997394	0.998554
ROC-AUC	0.999805	1.000000	0.999991	0.999976	0.999965
RMSE	0.042665	0.034208	0.032851	0.042251	0.042886
MAE	0.014461	0.014506	0.009785	0.010627	0.011536

Figure 27 shows a confusion matrix that was developed for use in performance evaluation. Here, the confusion matrix helps demonstrate the performance of the recommendation model when it comes to classifying actual recommendations and predicted recommendations. The recommendation model was able to classify 1,839 cases as 'Not Recommended' and 2,155 as 'Recommended'. The number of false positives was 2, and false negatives were 4; hence, an exceptionally high overall classification accuracy and reliable performance for predicting recommendations using the recommendation model were demonstrated.

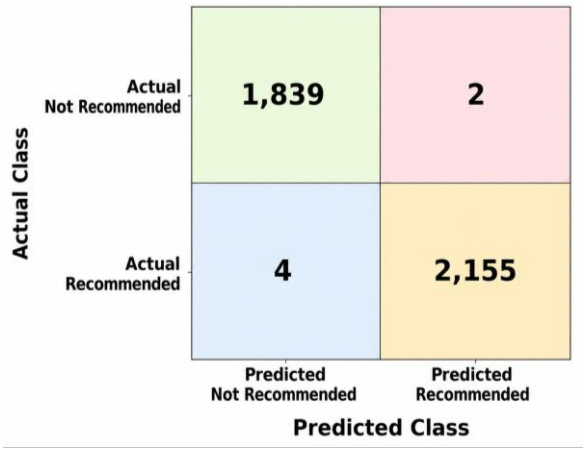

Figure 27: Confusion Matrix

The evaluation outcomes shown in Table 5 illustrate the effectiveness and trustworthiness of the proposed NSAR-T framework for generating personalized recommendations. The model racked up an Accuracy of 0.99, a Precision of 0.99, a recall of 0.99, and an F1-Score of 0.99, so overall classification looks extremely accurate. It also reported a Receiver Operating Characteristic (ROC)-AUC of 1.0, RMSE 0.036, plus MAE 0.011, which suggests the prediction errors stay very small. As for the ordering or ranking part, it was similarly strong: MAP got to 0.99, and NDCG@10 hit 1.0 exactly, meaning the recommendation list stays highly relevant and well ranked.

Table 5: Classification Metrics

Classification Metric	Metric Value
Accuracy	0.9985
Precision	0.9991
Recall	0.9981
F1-Score	0.9986
ROC-AUC	1.0000
RMSE	0.036
MAE	0.011
MAP	0.99
NDCG@10	1

4.1 Comparative Analysis

Table 6 compares the performance of several recommendation & sentiment analysis methods based on classification accuracy. The results show that the traditional methods, Multinomial Logistic Regression (MLR) and NB, achieved the lowest classification accuracy (68.0%), followed by the Support Vector Machine (SVM) with Linear Kernel (86.0%). In contrast, both of the Transformer-based machine learning methods surpassed those results, with BERT and Knowledge-Enhanced BERT (KE-BERT) having classification accuracy levels of 91.0% & 92.7%, respectively. The standard BERT has a classification accuracy of 91.63%. The NSAR-T outperformed all other methods, achieving the highest classification accuracy (99.85%), due primarily to its use of RoBERTa-based transformer modeling, ABSA, and seven levels of fine-grained sentiment classification, allowing for improved contextual understanding of sentiment, enhanced personalization quality, and more accurate recommendations.

Table 6: Comparative Analysis

Author	Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	MAE/ RMSE	Sentiment analysis

Promi et al., (2026) [58]	BERT-base-uncased Transformer model	7,567 Bangladeshi hotel reviews	79.45	80.14	79.95	79.71	NA	Involve
Irumporai et al., (2025) [59]	Knowledge-Enhaanced BERT (KE-BERT)	Social Media Sentiment Analysis Dataset	91.2	89.8	90.2	90.0	NA	Involve
Reddy et al., (2024) [39]	Bidirectional Gated Recurrent Units (Bi-GRU)	Amazon website reviews	96.9	96.2	NA	96.97	0.7869 / 0.5893	Involve
Solomenko and Symeon (2024) [60]	SVM BOW classifier	TripAdvisor website reviews	68	68	68	67	Na	Not involve
Namee et al., (2023) [61]	SVM, state-of-the-art	TripAdvisor website reviews	0.86	2.50	2.44	1.84	NA	Involve
This study	NSAR-T	200000 reviews TripAdvisor, Expedia website	99.85	99.91	99.81	99.86	0.03	Involve

5 Conclusion and Future Scope

Digital platforms and online review systems are growing rapidly, thus requiring a need for sentiment-based recommendation systems to provide a more personalized user experience. Because of the way that conventional recommendation approaches are created, they usually don't identify context or fine-grained emotional differences in reviews. Because of this, advanced intelligent recommendation solution frameworks need to be developed. The newly developed NSAR-T framework incorporates ABSA with a RoBERTa transformer model to create highly personalized recommendations using seven levels of fine-grained sentiment classification. The framework demonstrates the effectiveness of using the Primary dataset containing 200,000 web-scraped hotel reviews, the TripAdvisor dataset containing 878,561 hotel reviews, and the Expedia dataset containing ~37.6 million records. Key processes in the methodology for this study included preprocessing, aspect extraction, sentiment classification via RoBERTa, sentiment aggregation, feature engineering via PCA, cross-validation, and creating a Top-N recommendation set. The experimental results demonstrated strong performance with an accuracy of 99.85%, precision of 99.91%, recall of 99.81%, F1-score of 99.86%, RMSE of 0.036, MAP of 0.99, and NDCG@10 of 1, exceeding the performance levels of any of the current recommendation models. Future research on extending the existing framework should utilize multilingual transformer models, multimodal sentiment analysis, a real-time recommendation method, and the use of generative AI-based adaptive recommendation systems to advance the framework to provide improved scalability, contextual awareness, and dynamic personalization of recommendations.

Acknowledgement

We would like to express our sincere gratitude to all those who contributed to the successful completion of this research. We are particularly thankful for the guidance, support, and resources that were made available throughout the course of this work. The insights and encouragement we received played a vital role in shaping this study, and we truly appreciate the assistance provided at every stage.

References

1. Khrais, L.T., Gabbori, D., 2023. The effects of social media digital channels on marketing and expanding the industry of e-commerce within the digital world. *Periodicals of Engineering and Natural Sciences* 11(5), 64–75. https://doi.org/10.21533/pen.v11.i5.185

2. Elwarraki, O., Aammou, S., Jdidou, Y., Lahiassi, J., 2023. Toward the future of personalized learning: Emerging trends and challenges in recommendation systems. *ICERI2023 Proceedings*, 7226–7235. https://doi.org/10.21125/iceri.2023.1797
3. Nwanna, M., Offiong, E., Ogidan, T., Fagbohun, O., Ifaturoti, A., Fasogbon, O., 2025. AI-driven personalisation: Transforming user experience across mobile applications. *Journal of Artificial Intelligence, Machine Learning and Data Science* 3(1), 1930–1937. https://doi.org/10.51219/JAIMLD/maxwell-nwanna/425
4. Khalique, A., Rahmani, M.K.I., Saquib, M., Hussain, I., Muzaffar, A.W., Ahad, M.A., Nafis, M.T., Ahmad, M.W., 2022. A deterministic model for determining the degree of friendship based on mutual likings and recommendations on OTT platforms. *Computational Intelligence and Neuroscience* 2022(1), Article 9576468. https://doi.org/10.1155/2022/9576468
5. Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M., Luo, S., 2021. A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning (iJET)* 16(3), 274–306. https://doi.org/10.3991/ijet.v16i03.18851
6. Patoulia, A.A., Kiourtis, A., Mavrogiorgou, A., Kyriazis, D., 2023. A comparative study of collaborative filtering in product recommendation. *Emerging Science Journal* 7(1), 1–15. https://doi.org/10.28991/ESJ-2023-07-01-01
7. Akkaya, B., 2025. Current trends in recommender systems: A survey of approaches and future directions. *Computer Science* 10(1), 53–91. https://doi.org/10.53070/bbd.1652022
8. Fu, L., Ma, X., 2021. An improved recommendation method based on content filtering and collaborative filtering. *Complexity* 2021(1), Article 5589285. https://doi.org/10.1155/2021/5589285
9. Marcuzzo, M., Zangari, A., Albarelli, A., Gasparetto, A., 2022. Recommendation systems: An insight into current development and future research challenges. *IEEE Access* 10, 86578–86623. https://doi.org/10.1109/ACCESS.2022.3194536
10. Sarkar, S., 2025. Development of a hybrid recommendation system using collaborative filtering and content-based filtering techniques. *IRE Journals* 8(11), 2284–2292. https://www.irejournals.com/paper-details/1708607
11. Jadiga, S., 2025. Understanding the role of AI in personalized recommendation systems, applications, concepts, and algorithms. *International Journal of Computer Trends and Technology (IJCTT)* 73(1), 106–118. https://doi.org/10.14445/22312803/IJCTT-V73I1P113
12. Tegene, A., Liu, Q., Gan, Y., Dai, T., Leka, H., Ayenew, M., 2023. Deep learning and embedding-based latent factor model for collaborative recommender systems. *Applied Sciences* 13(2), Article 726. https://doi.org/10.3390/app13020726
13. Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., et al., 2023. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems* 1(1), 1–51. https://doi.org/10.1145/3568022
14. Sharma, K., Lee, Y.-C., Nambi, S., Salian, A., Shah, S., Kim, S.-W., Kumar, S., 2024. A survey of graph neural networks for social recommender systems. *ACM Computing Surveys* 56(10), 1–34. https://doi.org/10.1145/3661821
15. Chau, X.T.D., 2024. Analysing customer opinion, credibility, and interaction in user-generated content platforms. https://doi.org/10.25904/1912/5581
16. Li, S., Liu, F., Zhang, Y., Zhu, B., Zhu, H., Yu, Z., 2022. Text mining of user-generated content (UGC) for business applications in e-commerce: A systematic review. *Mathematics* 10(19), Article 3554. https://doi.org/10.3390/math10193554
17. Gheewala, S., Xu, S., Yeom, S., 2025. In-depth survey: Deep learning in recommender systems—exploring prediction and ranking models, datasets, feature analysis, and emerging trends. *Neural Computing and Applications* 37(17), 10875–10947. https://doi.org/10.1007/s00521-024-10866-z
18. Patel, C., 2026. Generative AI for personalized marketing and customer experience in e-commerce. *International Journal of Emerging Research in Engineering and Technology* 7(1), 12–19. https://doi.org/10.63282/3050-922X.IJERET-V7I1P103
19. Widayanti, R., Chakim, M.H.R., Lukita, C., Rahardja, U., Lutfiani, N., 2023. Improving recommender systems using hybrid techniques of collaborative filtering and content-based filtering. *Journal of Applied Data Sciences* 4(3), 289–302. https://doi.org/10.47738/jads.v4i3.115
20. Oprea, S.-V., Bâra, A., Ghinea, T., 2026. Recommender systems: Emerging trends from four decades of research using bibliometric analysis and transformer-based models. *Electronics* 15(4), Article 763. https://doi.org/10.3390/electronics15040763
21. Khan, Z.Y., Niu, Z., Sandiwarno, S., Prince, R., 2021. Deep learning techniques for rating prediction: A survey of the state-of-the-art. *Artificial Intelligence Review* 54(1). https://doi.org/10.1007/s10462-020-09892-9
22. Zhou, X., 2025. A brief review of basic deep learning models for recommendation systems. https://doi.org/10.5220/0013526300004619
23. Mohamed, H.A.I.I.M., 2020. *Deep learning models for research paper recommender systems*. Ph.D. dissertation. https://arcadia.sba.uniroma3.it/bitstream/2307/40817/1/PhD_Thesis%20-%20Hebatallah%20Mohamed.pdf
24. Li, P., Noah, S.A.M., Sarim, H.M., 2024. A survey on deep neural networks in collaborative filtering recommendation systems. *arXiv preprint* arXiv:2412.01378. https://doi.org/10.48550/arXiv.2412.01378
25. Ahmadian Yazdi, H., Seyyed Mahdavi, S.J., Ahmadian Yazdi, H., 2024. Dynamic educational recommender system based on an improved LSTM neural network. *Scientific Reports* 14(1), Article 4381. https://doi.org/10.1038/s41598-024-54729-y
26. de Souza Pereira Moreira, G., Rabhi, S., Lee, J.M., Ak, R., Oldridge, E., 2021. Transformers4Rec: Bridging the gap between NLP and sequential/session-based recommendation. In: *Proceedings of the 15th ACM Conference on Recommender Systems*. ACM, New York, pp. 143–153. https://doi.org/10.1145/3460231.3474255

27. Morales-Murillo, V.G., Pinto, D., Perez-Tellez, F., Rojas-Lopez, F., 2024. A transformer-based multi-domain recommender system for e-commerce. *International Journal of Combinatorial Optimization Problems and Informatics* 15(2), 95. https://doi.org/10.61467/2007.1558.2024.v15i2.465
28. Khan, H.U., Naz, A., Alarfaj, F.K., Almusallam, N., 2025. A transformer-based architecture for collaborative filtering modeling in personalized recommender systems. *Scientific Reports* 15(1), Article 24503. https://doi.org/10.1038/s41598-025-08931-1
29. Guo, X., 2024. Sentiment analysis based on RoBERTa for Amazon review: An empirical study on decision making. *arXiv preprint* arXiv:2411.00796. https://doi.org/10.48550/arXiv.2411.00796
30. Ayemowa, M.O., Ibrahim, R., Khan, M.M., 2024. Analysis of recommender system using generative artificial intelligence: A systematic literature review. *IEEE Access* 12, 87742–87766. https://doi.org/10.1109/ACCESS.2024.3416962.
31. Hong, Z., Wu, Y., Zhao, Z., Feng, S., Ma, J., Liu, J., Wei, T., 2025. Multi-objective recommendation in the era of generative AI: A survey of recent progress and future prospects. *arXiv preprint* arXiv:2506.16893. https://doi.org/10.48550/arXiv.2506.16893
32. Deldjoo, Y., He, Z., McAuley, J., Korikov, A., Sanner, S., Ramisa, A., Vidal, R., Sathiamoorthy, M., Kasirzadeh, A., Milano, S., 2024. A review of modern recommender systems using generative models (Gen-RecSys). In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM, New York, pp. 6448–6458. https://doi.org/10.1145/3637528.3671474
33. Deldjoo, Y., He, Z., McAuley, J., Korikov, A., Sanner, S., Ramisa, A., Vidal, R., et al., 2026. Recommendation with generative models. *Foundations and Trends® in Information Retrieval* 20(1–2), 1–176. https://doi.org/10.1108/FTINR-06-2025-0109
34. Bharadwaj, L., 2023. Sentiment analysis in online product reviews: Mining customer opinions for sentiment classification. *International Journal of Multidisciplinary Research* 5(5). https://pdfs.semanticscholar.org/fdca/618ff12355dad6371f26a7d17018a63f8793.pdf
35. Paramita, A.S., Jusak, 2025. Fine-grained sentiment analysis approach on customer reviews based on aspect-level emotion detection. *Journal of Applied Data Sciences* 6(3), 2235–2247. https://doi.org/10.47738/jads.v6i3.964
36. Truong, V., 2025. Examining the sentiment and emotional differences in product and service reviews: The moderating role of culture. *arXiv preprint* arXiv:2507.21057. https://doi.org/10.48550/arXiv.2507.21057
37. Kotei, E., Thirunavukarasu, R., 2023. A systematic review of transformer-based pre-trained language models through self-supervised learning. *Information* 14(3), Article 187. https://doi.org/10.3390/info14030187
38. Acheampong, F.A., Nunoo-Mensah, H., Chen, W., 2021. Transformer models for text-based emotion detection: A review of BERT-based approaches. *Artificial Intelligence Review* 54(8), 5789–5829. https://doi.org/10.1007/s10462-021-09958-2
39. Reddy, B., Kumar, R., 2024. A fusion model for personalized adaptive multi-product recommendation system using transfer learning and bi-GRU. *Computers, Materials & Continua* 81(3), Article 4081. https://doi.org/10.32604/cmc.2024.057071.
40. Shrestha, D., Wenan, T., Shrestha, D., Rajkarnikar, N., Jeong, S.-R., 2024. Personalized tourist recommender system: A data-driven and machine-learning approach. *Computation* 12(3), Article 59. https://doi.org/10.3390/computation12030059
41. Abbasi, F., Khadivar, A., Yazdinejad, M., 2019. A grouping hotel recommender system based on deep learning and sentiment analysis. *Journal of Information Technology Management* 11(2), 59–78. https://doi.org/10.22059/jitm.2019.289271.2402
42. Afroze, S., Hossain, M.R., Hoque, M.M., Siddique, N., 2026. MulMoSent: Multimodal sentiment analysis for a low-resource language using textual-visual cross-attention and fusion. *Information Fusion*, Article 104129. https://doi.org/10.1016/j.inffus.2026.104129
43. Pan, Z., Liang, Z., 2025. Gated dynamic local-global attention for sentiment analysis: Enhancing context awareness in short texts. *Authorea Preprints*. https://doi.org/10.36227/techrxiv.174495340.03183932/v1
44. Çaylak, P.Ç., Kayakuş, M., Eksili, N., Açikgöz, F.Y., Coşkun, A.E., Ichimov, M.A.M., Moiceanu, G., 2024. Analysing online reviews consumers' experiences of mobile travel applications with sentiment analysis and topic modelling: The example of Booking and Expedia. *Applied Sciences* 14(24), Article 11800. https://doi.org/10.3390/app142411800
45. Kaur, G., Sharma, A., 2023. A deep learning-based model using a hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data* 10(1), Article 5. https://doi.org/10.1186/s40537-022-00680-6
46. Santhi, S., Thavasi, S., Umakanth, N., 2022. Suggesting hotels through reviews using sentiment analysis. In: *Ambient Communications and Computer Systems: Proceedings of RACCCS 2021*. Springer Nature Singapore, Singapore, pp. 595–605. https://doi.org/10.1007/978-981-16-7952-0_57
47. Wang, H.-C., Justitia, A., Wang, C.-W., 2024. AsCDPR: A novel framework for ratings and personalized preference hotel recommendation using cross-domain and aspect-based features. *Data Technologies and Applications* 58(2), 293–317. https://doi.org/10.1108/DTA-03-2023-0101
48. You, T.-H., Tao, L.-L., Cambria, E., 2023. A hotel ranking model through online reviews with aspect-based sentiment analysis. *International Journal of Information Technology & Decision Making* 22(1), 89–113. https://doi.org/10.1142/S0219622022500626
49. Yang, L., 2026. Research on user preference modeling and data enhancement based on generative adversarial networks in precision recommendation for e-commerce platforms. *International Journal of Computer Information Systems and Industrial Management Applications* 18, 15–15. https://doi.org/10.70917/ijcisim-2026-0134

50. Rizvi, A., Thamindu, N., Adhikari, A.M.N.H., Senevirathna, W.P.U., Kasthurirathna, D., Abeywardhana, L., 2025. Enhancing multilingual sentiment analysis with explainability for Sinhala, English, and code-mixed content. *arXiv preprint* arXiv:2504.13545. https://doi.org/10.48550/arXiv.2504.13545
51. Ray, B., Garain, A., Sarkar, R., 2021. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews. *Applied Soft Computing* 98, Article 106935. https://doi.org/10.1016/j.asoc.2020.106935
52. Chong, W.K., Ng, H., Yap, T.T.V., Tan, I.K.T., Goh, V.T., Wong, L.K., Cher, D.T., 2026. Multitask learning and BERT embedding: A comprehensive approach to subjectivity detection and aspect-based sentiment analysis. *International Journal of Computational Intelligence Systems*. https://doi.org/10.1007/s44196-026-01153-x
53. Narayanaswamy, G.R., 2021. *Exploiting BERT and RoBERTa to improve performance for aspect-based sentiment analysis*. https://doi.org/10.21427/3w9n-we77
54. Kumar, B.V., Sadanandam, M., 2024. A fusion architecture of BERT and RoBERTa for enhanced performance of sentiment analysis of social media platforms. *International Journal of Computing and Digital Systems* 15(1), 51–66. https://doi.org/10.12785/ijcds/150105
55. Nanga, S., Bawah, A.T., Acquaye, B.A., Billa, M.-I., Baeta, F.D., Odai, N.A., Obeng, S.K., Nsiah, A.D., 2021. Review of dimension reduction methods. *Journal of Data Analysis and Information Processing* 9(3), 189–231. https://doi.org/10.4236/jdaip.2021.93013
56. Promi, S.A., Absar, N., Hafiz, M.S., n.d. Deep learning-driven aspect-based sentiment analysis of Bangladeshi hotel reviews. ResearchGate. https://www.researchgate.net/publication/401583196_Deep_Learning-Driven_Aspect-Based_Sentiment_Analysis_of_Bangladeshi_Hotel_Reviews
57. Irumporai, A., Anitha, G., 2025. Knowledge-enhanced BERT for aspect-based sentiment analysis of tourism destinations using social media data. *Engineering, Technology & Applied Science Research* 15(6), 30097–30102. https://doi.org/10.48084/etasr.13139
58. Solomenko, D., Charalabides, S., 2024. *Aspect extraction for hotel reviews: A hybrid approach*. https://doi.org/10.13140/RG.2.2.15692.42881
59. Namee, K., Polpinij, J., Luaphol, B., 2023. A hybrid approach for aspect-based sentiment analysis: A case study of hotel reviews. *Current Applied Science and Technology*, Article 10-55003. https://doi.org/10.55003/cast.2022.02.23.008